tesol QUARTERLY
tesol international association

How Does Having a Good Ear Promote Instructed Second Language Pronunciation Development? Roles of Domain-General Auditory Processing in Choral Repetition Training

YUJIE SHAO AND KAZUYA SAITO
University College London
London, United Kingdom
Chengdu No.7 High School
Sichuan, China

ADAM TIERNEY
Birkbeck, University of London
London, United Kingdom

Abstract

Growing evidence suggests that auditory processing ability may a crucial determinant of language learning, including adult second language (L2) speech learning. The current study tested 47 Chinese English-as-a-Foreign-Language students to examine the extent to which two types of auditory processing, i.e., perceptual acuity and audio-motor integration, related to improvements in the comprehensibility and nativelikeness of L2 speech following two weeks of choral repetition training (i.e., shadowing). All participants' pronunciation proficiency became significantly more comprehensible over time, and the degree of improvement in the nativelikeness of pronunciation was tied to the ability to remember and reproduce sounds (i.e., audio-motor integration). The findings suggest that robust auditory-motor integration may play a key role in the acquisition of advanced-level L2 pronunciation proficiency (i.e., comprehensible and nativelike speech).
doi: 10.1002/tesq.3120

Individuals differ widely in terms of how they process domain-general acoustic information such as pitch, duration, and spectral patterns (Kidd, Watson, & Gygi, 2007). On a broader level, this perceptual

ability, collectively termed as domain-general auditory processing, comprises two different constructs—(a) the extent to which they can hear very subtle acoustic details of sounds (i.e., auditory acuity) and (b) how well they can convert perceived information into motor action (i.e., audio-motor integration). An emerging paradigm in the field of second language (L2) acquisition suggests that these differences in auditory processing can modulate adult language learning trajectories: individuals with more precise auditory processing abilities may be better able to utilize every input opportunity, resulting in greater gains in the long run (i.e., the auditory precision hypothesis; Mueller, Friederici, & Männel, 2012; Kachlicka, Saito, & Tierney, 2019; for a comprehensive overview, see Saito & Tierney, forthcoming).

Following the auditory account of L2 acquisition, prior cross-sectional research has shown that participants' auditory processing abilities are tied to the outcomes of successful L2 speech learning in both naturalistic (e.g., Saito, Kachlicka, Sun, & Tierney, 2020) and classroom settings (e.g., Saito, Suzukida, Tran, & Tierney, 2021). In the context of 47 Chinese English-as-a-Foreign-Language (EFL) learners, the current longitudinal study tested whether, to what degree, and how different constructs of auditory processing ability (acuity, integration) could predict learners' multi-dimensional improvement in L2 pronunciation (comprehensibility, nativelikeness) following two weeks of choral repetition training (i.e., shadowing).

BACKGROUND

Assessing and Teaching L2 Pronunciation

Over the past 50 years, much attention has been given to devising optimal methods for assessing and teaching L2 pronunciation (i.e., the accurate and fluent articulation of new vocalic and consonantal sounds with adequate and varied stress and intonation patterns). While many teachers, students, and textbook developers have emphasized the attainment of nativelike pronunciation skills as an *ideal* goal (e.g., Derwing, 2003 for Canadian ESL classrooms), research has convincingly shown that adult L2 pronunciation is generally foreign-accented due to the influence of fully established L1 phonetic systems (see Flege & Bohn, 2021 for a theoretical account of the complexities underlying L1-L2 interactions). Therefore, many scholars have argued that what matters most for communicative success is the degree to which an L2 user's speech is comprehensible, intelligible, and communicatively adequate (Levis, 2018).

To date, scholars have conceptualized, discussed, and operationalized L2 pronunciation proficiency from two different angles: comprehensibility and accentedness.[1] Comprehensibility concerns listeners' ease of understanding, while accentedness relates to phonological nativelikeness (Derwing & Munro, 2013). Both constructs are typically measured using listeners' intuitive judgments of L2 speech. During L2 comprehensibility judgments, listeners have been shown to attend to a variety of dimensions of L2 speech, including phonological accuracy, speaking fluency (Trofimovich & Isaacs, 2012), varied prosody (Kang, Rubin, & Pickering, 2010), and the varied, sophisticated, and contextually appropriate use of lexicogrammar (Appel, Trofimovich, Saito, Isaacs, & Webb, 2019). In other words, when rating L2 speech, listeners aim to collect as much linguistic information as possible to grasp an overall picture of what the speaker intends to convey (Saito, Trofimovich, & Isaacs, 2017). When it comes to L2 accentedness judgments, however, it has been shown that listeners preferentially attend to a speaker's degree of phonetic refinement, especially at the segmental level, rather than factors related to understanding speech content (Trofimovich & Isaacs, 2012).

Research in classroom (Nagle, 2018) and naturalistic settings (Derwing & Munro, 2013) has shown that L2 learners can substantially enhance their comprehensibility if they use the target language on a regular basis. In addition, it has been shown that L2 comprehensibility is particularly amenable to improvement via explicit instruction (Saito & Plonsky, 2019). On the other hand, it has been shown that L2 accentedness is resistant to change even if learners engage in explicit instruction (Derwing, Munro, Foote, Waugh, & Fleming, 2014). This is arguably because the primary correlate of foreign accentedness—segmental refinement—requires an extensive amount of L2 immersion experience from an early age (Flege & Bohn, 2021). Furthermore, the acquisition of nativelike L2 phonology appears to be limited to certain individuals with certain perceptual-cognitive abilities (e.g., He et al., 2013 for phonemic coding).

Research on L2 speech development suggests that comprehensibility, rather than accentedness, may be a more ecologically valid goal for most adult L2 learners (Isaacs, Trofimovich, & Foote, 2018). While no learner should be discouraged from attaining nativelike pronunciation

[1] "Pronunciation proficiency" is a very difficult phenomenon to define. Following Saito and Plonsky's (2019) model of instructed L2 pronunciation proficiency, listener ratings of comprehensibility and accentedness were used as one index of global L2 pronunciation proficiency in the current manuscript. To support this, Isaacs, Trofimovich, Yu, and Muñoz Chereau (2015) showed that L2 speakers' speaking proficiency scores (measured via IELTS Speaking Scale) were significantly correlated with the comprehensibility and phonological nativelikeness of their spontaneous speech (r =.509 and.585, respectively).

proficiency, it is important that teachers inform students that (a) reducing one's foreign accent is a relatively difficult task after puberty; (b) that even foreign-accented speech can be highly comprehensible; and (c) that such comprehensibility can be enhanced via practice (for a meta-analysis, Saito, 2021).

In this current paper, advanced L2 speech proficiency is defined as not only comprehensible but also nativelike pronunciation.[2]

Shadowing, Tracking, and Choral Repetition Training

While various teaching approaches have been applied to L2 pronunciation, growing attention has been given to the pedagogical potential of tracking and shadowing activities (for a comprehensive overview, Hamada, 2018). According to Celce-Murcia, Brinton, and Goodwin (2010), tracking is where learners listen to native speakers either face-to-face or remotely on television, radio, or audiotape while following a transcript or subtitles, and simultaneously reproduce what they hear. Shadowing is similar to tracking in that learners listen and repeat what they hear with a slight delay and can pause the recording at times if necessary.

The common element between tracking and shadowing is the choral repetition of model speech. While this concept is reminiscent of the audio-lingual training methods of years past, it remains popular in many foreign language classrooms all over the world (e.g., Saito & van Poeteren, 2012 for the results of experienced EFL teachers in Japan) and is an integral component of digital speaking training materials (e.g., Rosetta Stone; see Lord, 2015). The primary objective of the method is to help L2 learners increase their control over what they have already heard, learned, and remembered.

There are several reasons why tracking, shadowing, and repetition activities are believed to facilitate L2 pronunciation development. First,

[2] In the current investigation, the comprehensibility and accentedness aspects of L2 pronunciation proficiency were considered. Whereas such scores are based on listeners' subjective ratings, Derwing and Munro have argued that what is ultimately important for communicative success is *intelligibility*, i.e., interlocutors' actual understanding of intended message. In the current investigation, we do not intend to introduce/discuss the literature on intelligibility. This is because the method and interpretation of intelligibility has been substantially different and mixed among scholars. For example, a wide range of methods have been used, such as transcription, comprehension questions, scaler ratings, and reaction time instruments (for a comprehensive review on methodological fuzziness in L2 intelligibility research, see Isaacs, 2008). Thus, we decided not to include the notion of intelligibility in the sections of literature review, research design, and future directions, because there is no methodological consensus on intelligibility in the field of L2 pronunciation research, and intelligibility was not measured in the current study. More research is needed to further explore the mechanisms underlying intelligibility and the development of more empirically robust methods.

these activities provide learners with additional input opportunities that learners otherwise lack in foreign language classrooms (Muñoz, 2014). Second, learners are encouraged to notice and fill in the gap between their current proficiency levels and targetlike performance by comparing their output to the input they hear; this alignment process is a key factor of successful L2 speech learning (Trofimovich, Isaacs, Kennedy, Saito, & Crowther, 2016). Third, learners can focus on the imitation of nativelike pronunciation forms which is considered to be an important skill for successful L2 speech learning (Flege, Munro, & MacKay, 1995). Fourth, systematic repetition activities can reduce the frequency of dysfluencies and promote the automatization of L2 speech production (Suzuki, 2021).

Considerable attention has been given to examining how and whether shadowing and tracking can be used to improve various dimensions of L2 pronunciation proficiency, such as segmental accuracy (Hamada, 2018), word and sentence stress (Martinsen, Montgomery, & Willardson, 2017), intonation (Hsieh, Dong, & Wang, 2013), intelligibility (Martinsen et al., 2017), and comprehensibility (Hamada, 2018). However, very few of these studies have examined L2 pronunciation development using spontaneous speaking tasks (Martinsen et al., 2017).

Individual Differences in Instructed L2 Speech Learning

Studies conducted in the classroom setting have pointed to several different types of explicit instruction which can help improve L2 pronunciation proficiency (e.g., Sakai & Moorman, 2018 for high variability perception training; Derwing et al., 2014 for prosody-based production training; Suzuki, 2021 for fluency training). Although there is a consensus that the provision of explicit instruction (including shadowing and tracking) facilitates adult L2 pronunciation learning, the effect sizes of explicit instruction are relatively small-to-medium (e.g., d =.078; Saito & Plonsky, 2019). Here, it is important to note that the outcomes of training are subject to a great deal of individual variation. Even if two participants spend the same amount of time practicing a target language on a daily basis in the same setting, their learning patterns may differ to a great degree (Doughty, 2019).

Though few in number, some studies have begun to examine which individual difference factors can help learners gain the most from explicit pronunciation instruction. Evidence collected to date suggests that more advanced L2 learners have greater levels of phonological awareness (Venkatagiri & Levis, 2007), phonemic coding ability (He

et al., 2013), and working memory capacity (Darcy, Park, & Yang, 2015). In the current investigation, we would like to introduce domain-general auditory processing as a potential determinant of language learning that may be germane to adult L2 speech learning (Saito & Tierney, forthcoming). Then, we hypothesize that those with more precise and more robust auditory processing abilities will benefit more from shadowing training.

In the following section, we will provide a detailed literature review on (a) what comprises auditory processing (i.e., auditory acuity for input encoding; audio-motor integration for linking auditory input with motor output), (b) what prior research has revealed about the roles of two different auditory abilities in L2 speech learning (i.e., acuity for the attainment of naturalistic L2 speech learning; integration for the optimization of production-based L2 practice in classroom settings), and (c) why one of the components in particular (i.e., integration rather than acuity) could be predictive of certain individuals who can benefit most from shadowing training and achieve more advanced L2 pronunciation. Following the comprehensibility-accentedness model of L2 pronunciation proficiency (Derwing & Munro, 2013), such advanced L2 pronunciation proficiency is defined as comprehensible *and* nativelike in the current investigation.

Domain-General Auditory Processing in L1 and L2 Acquisition

Oral input is important during every phase of language learning. When receiving linguistic input, learners must encode time and frequency patterns (i.e., auditory acuity) and subsequently remember and integrate them into action (i.e., audio-motor integration). These auditory processing skills have been hypothesized to anchor various dimensions of language learning (i.e., Auditory Precision Hypothesis; Goswami, 2015). For instance, auditory processing has been shown to be relevant to a variety of linguistic skills, including detection of different speech contrasts (Werker & Tees, 1999), prosodic patterns (De Pijper & Sanderman, 1994), lexical boundaries (Cutler & Butterfield, 1992), syntactic constructions (Marslen-Wilson, Tyler, Warren, Grenier, & Lee, 1992), and discourse structures (Wang, Li, & Yang, 2014).

To date, there is evidence among L1 acquisition studies showing a significant association between individual differences in auditory processing and a range of language acquisition phenomena, such as literacy development (White-Schwoch et al., 2015) and reading competency (Boets et al., 2011). Based on these findings, some researchers have suggested that auditory processing ability could be

used as a diagnostic tool for certain types of language delay or impairment, such as dyslexia (Hornickel & Kraus, 2013).

Building on the L1 acquisition literature, some scholars have suggested that auditory processing may play a significant role in adult L2 speech learning as well (Kachlicka et al., 2019; Mueller et al., 2012; Saito & Tierney, forthcoming). Unlike L1 acquisition, which is free of the influence of prior language learning experience, adult L2 learners filter input through their L1 acoustic representations. Because of this, they could face a tremendous amount of difficulty identifying and discriminating new acoustic dimensions that they do not usually use when detecting L1 phonological contrasts (e.g., F3 discrimination for the English /r/ and /l/ contrast for Japanese learners; Saito, 2013a).

On a broad level, auditory processing can be divided into two distinctive abilities, auditory acuity and audio-motor integration. Auditory acuity concerns one's ability to hear a very subtle difference in frequency and time characteristics of sounds. This ability has been measured via a psychoacoustic task where participants discriminate synthesized sounds which differ in one particular acoustic domain (e.g., formant, pitch, duration). Audio-motor integration relates to one's ability to link auditory input with motor action. This ability has been measured via reproduction tasks where participants replicate sets of melodic and rhythmic sequences (for comprehensive overviews of the auditory precision hypothesis – L2, see Saito, Suzukida, et al., 2021; Saito & Tierney, forthcoming).

Recent research suggests that these two different types of auditory processing abilities (auditory acuity and audio-motor integration) can uniquely predict successful L2 speech learning in immersion and classroom contexts. In the immersion settings, advanced L2 speakers with extensive length of residence profiles were shown to also have precise auditory acuity (Zheng, Saito, & Tierney, 2021) and audio-motor integration abilities (Kachlicka et al., 2019; Saito, Kachlicka, et al., 2020; Sun, Saito, & Tierney, 2021). By comparison, research conducted in the foreign language classroom setting has shown that L2 speakers with relatively advanced L2 speech profiles are more likely to have higher audio-motor integration abilities (Saito, Suzukida, et al., 2021), but not auditory acuity (cf. Saito, Sun, et al., 2020; Sun et al., 2021). This is arguably because (a) they have limited exposure to communicatively authentic, aural input; (b) most of the input they receive is delivered by teachers who are nonnative speakers; and (c) their L2 use is restricted to only a few hours of traditional production-based instruction per week (e.g., grammar translation, audio-lingual method) (Muñoz, 2014; Saito, Suzukida, et al., 2021).

In summary, whereas auditory processing matters for L2 speech learning, two subcomponents of auditory processing—i.e., acuity and integration—may differentially facilitate the attainment of advanced

L2 speech proficiency. Acuity is a crucial variable for the attainment of high-level L2 speech proficiency in naturalistic settings wherein learners access a number of communicatively authentic input opportunities. If one has more precise acuity ability, it will help further encode very subtle details of acoustic signals and refine the quality of their own acoustic representations. Contrastively, integration can serve as a key factor of successful L2 speech learning especially in classroom settings where learners lack enough input but are encouraged to produce motor output based on a limited amount of input. Robust integration skills could assist with the rapid capturing of broad acoustic information and subsequent prompt conversion to motor action.

It is noteworthy that most studies carried out to date have been cross-sectional in nature. These studies have yet to explore how the relationship between auditory and L2 speech abilities change over time. To our knowledge, very few studies have examined to what degree auditory perception skills can help L2 learners process input more effectively and efficiently over time. There is some emerging evidence that L2 learners with more precise acuity abilities benefited more from perception-based training (e.g., Qin, Zhang, & Wang, 2021) and that those with more robust integration abilities demonstrated larger gains from production-based training (Li & DeKeyser, 2017). However, no studies have examined the relative weights of acuity and integration in L2 speech learning. The current study took an exploratory approach toward examining how two different auditory processing skills—i.e., acuity vs. integration—could differentially predict L2 pronunciation learning gains in the context of traditional repetition-based pronunciation training (i.e., shadowing).

Current Study, Research Questions, and Predictions

The current study set out to provide the *longitudinal* evidence regarding the relationship between auditory processing and L2 speech learning. Here, we followed the definition of "longitudinal data" widely used in Applied Psychology which refers to data collected via "repeated measures of (at least twice) the same variables gathered from the same study participants over time" (Dormann & Guthier, 2019, p. 158). Some scholars have attempted to define what characterizes longitudinal research in the field of L2 acquisition, teaching, and pronunciation. Whereas cross-sectional design focuses on L2 learners' linguistic behaviors at a single data collection point, longitudinal research focuses on how their language develops over time thanks to a range of independent variables (e.g., training and immersion experience; for a comprehensive review, see Nagle, 2021). Following the

guidelines initially set by Ortega and Iberri-Shea (2005) and recently extended by Saito, Suzuki, Oyama, and Akiyama (2021), the study could be considered as "longitudinal" in nature, as it meets the three crucial conditions of such research design: (1) multiple sessions (the participants participated in 12 training sessions); (2) multiple data collection points (the participants' speech development was measured via pre- and post-tests); and (3) multiple types of analyses (the outcome measures comprised different types of elicitation tasks and listener judgments). The study was guided by two research questions:

1. The first question examined the extent to which the provision of choral repetition training (i.e., tracking and shadowing) could facilitate improvement on two different dimensions of L2 pronunciation proficiency—comprehensibility and accentedness.
2. The second question examined the extent to which auditory processing (auditory acuity vs. audio-motor integration) was associated with enhanced comprehensibility and reduced accentedness.

As for R1, we predicted that tracking and shadowing would lead to significant improvements in both comprehensibility and accentedness, as shown in Hamada (2018). As for R2, we predicted that those with greater integration (rather than acuity) ability would realize the most benefit. This is because these learners are assumed to be more capable of incorporating auditory information into motor action—a function which plays a vital role in accessing a new language more smoothly, quickly, and effortlessly. Moreover, shadowing requires listeners to rapidly encode auditory patterns extending across multiple segments and formulate the appropriate motor sequence for reproducing these patterns, a process that is directly targeted by the auditory-motor integration tests (Saito, Suzukida, et al., 2021). Finally, it was predicted that the aptitude-acquisition link would be observed most clearly for the most difficult dimension of L2 speech learning, i.e., participants' accentedness reduction at a spontaneous speech level.

METHOD

Participants

L2 learners. A total of 47 Chinese high school EFL learners (22 males and 25 females) participated in the study. All of them were third year students between 17 to 18 years of age. The high school is located in a suburban district in Chengdu, Sichuan. All the participants engaged in several hours of EFL instruction per week taught by Chinese EFL teachers. The content of EFL lessons was based on a

mixture of grammar translation, choral repetition, and meaning-oriented speaking and listening activities. An electronic flyer was disseminated at the high school, and interested participants contacted the investigator and participated in the project as an extracurricular EFL activity.

Although they had received an extensive amount of EFL experience prior to participation in the study (e.g., 6+ years), none of them had received any formal English pronunciation training nor engaged in immersion experience. The participants were divided into an Experimental group ($n = 37$) and a Control group ($n = 10$). Because the main focus of the study lay in the analyses of the individual differences, the number of the participants in the Experimental group was much larger than that of the Control group. All participants reported having normal hearing.

Raters. A total of five linguistically trained Chinese coders (1 male, 4 females; ages 23 - 27, $M = 24.8$) were recruited in London to act as speech raters. All of them were graduate students in applied linguistics, and had extensive knowledge of, and experience with, L2 speech analysis. All raters reported relatively high IELTS scores ($M = 7.5$, $SD = 0.7$), which classifies them as Advanced Users of English (C1, C2) according to CEFR benchmarks (Council of Europe, 2001). Following the evidence that highly advanced L2 and native speaking listeners make similar comprehensibility and accentedness judgments (Derwing & Munro, 2013), we did not consider the rater backgrounds (linguistically trained Chinese listeners) as a confounding factor and we did not make any attempts to recruit native speaking listeners with a view to comparing L2 and L1 listener ratings. Our methodological justification corresponds to the detailed analyses of L1 and L2 raters' *similar* comprehensibility and accentedness judgments in Saito's (2021) research synthesis. We obtained relatively high inter-rater reliability (α =.88 for comprehensibility; α =.85 for accentedness).

Design. Figure 1 visually summarizes the design of the study. All participants completed the pre-tests in Week 1. In Weeks 2 and 3, participants in the experimental group took the auditory processing tests (acuity, integration) and participated in 12 sessions of fluency training, while participants in the control group participated in 12 sessions of vocabulary, grammar, and writing practice. Finally, all participants completed immediate post-tests in Week 4. Because the same materials were used for the pre- and post-tests, the inclusion of the control group was crucial to control for any test-retest effects. In this way, it was possible to identify whether the gains from L2 speech

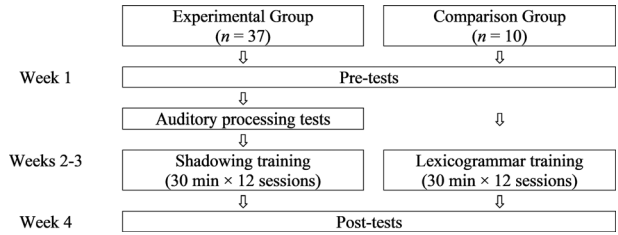

FIGURE 1. Summary of research design.

training were linked to individual differences in acuity and integration abilities.

Shadowing Training (Experimental Group)

As a part of their regular high school EFL activities, participants in the experimental group participated in a total of 12 shadowing training sessions over the course of two weeks (30 minutes × 12 sessions = 6 hours). In each session, the participants practiced shadowing using a popular application called *English Fun Dubbing* (EFD), which was downloaded onto iPads borrowed from the school. The EFD app presents short segments of videos along with line-by-line transcripts of the content. The app allows users to watch the videos while reading the scripts, and then imitate and record what they hear. To avoid any confusion about the training procedure, all instructions were delivered in Chinese by the researcher. The training sessions were operationalized as follows:

1. Dubbing materials: A total of three videos were selected from the CBeebies Bedtime Story series produced by the BBC: *Follow the Swallow*, *Mr Big*, and *What Friends Do Best* (for all scripts used in the current study, see Supporting Information-A). The videos were chosen to conform to the nature of the pre-/post-tests (i.e., spontaneous picture description), where participants were asked to describe a four-frame picture with a consistent storyline (for more details, see below the Outcome Measures section). All three videos were practiced during each session.
2. Instructional procedure: All the participants in the experimental group did the dubbing activities together in a single classroom at the high school. At each session, their EFL teacher and the first author instructed, guided, and monitored the participants. The participants were asked to do the assigned shadowing tasks with their iPad by using headphones. Whenever the

participants had questions about the procedure, the EFL teacher and the investigator provided individual support. To ensure that each participant had spent the equal amount of time on the assigned task, they were allowed to engage in the shadowing activities only during the training session. The content of shadowing was selected by the investigator for each session so that all the participants worked on the same materials. To avoid the impact of peers' noise on their performance and recording, the participants were scattered across the classroom and along the corridor. We acknowledge that the background noise could not be completely avoided due to space constraint. However, efforts were made to minimize the background noise. As a result, participants' shadowing practice was clearly recorded on their own device. Each participants' performance was checked by the investigator. Participants were first instructed to watch the videos while reading the scripts, and to repeat what they were hearing. During each video, the participants were able to press a button labeled "Start Dubbing" to begin recording their shadowing on a sentence-by-sentence basis. Participants were allowed to listen to the original audio, compare it with their shadowing performance, and re-dub the sentences as many times as they wanted to.

3. Publish and share: After dubbing the entire video, participants were able to preview their recordings and return to the dubbing interface to re-dub any sentences they desired. They then saved their dubbing recordings and shared them with their classmates via the app. No feedback was provided on any of the recordings. The participants were explicitly asked to practice the shadowing activities via EFD as much as possible within each training session (30 minutes). All the participants' recordings were saved by the investigator.

This form of shadowing training is similar to the treatment provided in many of the previous studies reviewed earlier (e.g., Hamada, 2018). However, the affordances of mobile technology allowed the participants to practice the speech shadowing at their own pace; that is, they could repeat each sentence as many times as they wanted to. Although the training was tailored to individuals' proficiency levels, the participants shadowed all the sentences in each video at least three times.

Lexicogrammar Training (Control Group)

Participants in the control group spent a similar amount of time (30 minutes × 12 sessions) practicing various aspects of English (e.g.,

vocabulary, grammar, writing) with the researcher over a period of two weeks. Most of the materials comprised follow-up activities based on what they had practiced in their weekly EFL lessons. No speaking practice or training was provided during these sessions. The purpose of the control group was to show the presence/absence of any test-retest effects as the same materials were used in pre- and post-tests.

Speech Outcome Measures

Pre/post-test materials. Both controlled and spontaneous speaking tasks were used to elicit L2 speech performance. For the controlled speaking tasks, participants were asked to read a total of six sentences that were randomly taken from the tracking and shadowing training materials without any preparation time:

1. It was covered in white blossom (*Follow a Swallow*).
2. The jumpy dolphin swam and leapt and dived (*Follow a Swallow*).
3. He was so big that anywhere he went, all everyone saw was someone big and scary (*Mr Big*).
4. It looked all alone, just like him, so he bought it and took it home (*Mr Big*).
5. Everything looked much the same as it had done a week ago, except Winston, who looked very miserable (*What Friends Do Best*).
6. Some of the pieces were very tiny and difficult to find (*What Friends Do Best*).

For the spontaneous speaking task, participants were asked to describe a four-frame picture story under time pressure. Two different versions of the pictures (Version A & B) were chosen from the EIKEN English Test (Pre-1 Level) (EIKEN, 2016). These were counterbalanced across group (experimental/control) and time (pre-/post-test). Version A was about an elderly couple who lived far away from the nearest supermarket, while Version B was about a girl who wanted to buy a smartphone. Two versions were chosen to avoid producing any confounding effects from using different topics. Students were given one minute to prepare their speech and two minutes to narrate the story. To avoid false starts, the first sentence of the narration task was provided for participants (see Supporting Information-B).

Procedure. All speech samples were recorded in a quiet office in the school using the audio recorder on an SM-G9750 mobile device, set to a 48 kHz sampling rate. To reduce rater fatigue, the first

30 seconds of the speech samples were cut and saved as MP3 files. In total, 188 speech samples (47 participants × 2 tasks × 2 testing points) were collected for rating.

Comprehensibility and Accentedness Judgments

Procedure. All speech samples were rated by five advanced L2 English users (Derwing & Munro, 2013). Due to the ongoing pandemic, the speech samples and rating guidelines were provided to the raters via an online cloud system. In addition, all rating sessions took place individually using a video-conferencing tool. First, the raters were given detailed explanations of comprehensibility (ease of understanding) and accentedness (linguistic nativelikeness) to ensure that they fully understood each construct. They then assessed the speech samples using a 9-point scale for comprehensibility (*1 = difficult to understand, 9 = easy to understand*) and accentedness (*1 = heavily accented, 9 = no accent*) on their own computers. The controlled speech samples were assessed before the spontaneous speech samples.

Reliability. Relatively high Cronbach's alpha levels were reported for the raters' comprehensibility ($\alpha = .88$) and accentedness ($\alpha = .85$) scores. The raters' scores were averaged for each construct for further analysis.

Auditory Processing Measures

Two types of auditory processing tests were used to assess auditory acuity and audio-motor integration: discrimination tests and reproduction tests (Moore, 2012). These were delivered via the online platform of psychology experiment named GORILLA (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020). In the first week, participants in the experimental group ($n = 37$) were invited to take the auditory processing tests individually using a laptop and a set of earphones in a quiet room. They first received detailed instructions about the tests before completing them in the following sequence: rhythm reproduction (integration), melody reproduction (integration), formant discrimination (acuity), pitch discrimination (acuity), and duration discrimination (acuity).

Auditory-motor integration (rhythm, melody). Based on the procedure and materials developed by Tierney, White-Schwoch, MacLean, and Kraus (2017), Saito, Suzuki, et al. (2021), and Sun et al. (2021),

two different audio-motor integration tasks were implemented—rhythm and melody reproduction.

- **Rhythm Reproduction**: As for stimuli, ten rhythmic patterns were created and presented to the participants (3.2 seconds per sample). These rhythms were taken from the rhythmic patterns used in Povel and Essens (1985). Each rhythm consisted of a series of 16 200-ms segments. In segments containing a drum hit, the first 150 ms of the segment contained a conga drum hit obtained from freesound.org. In segments containing a rest, no sound was present throughout the segment. In each trial, after listening to the same stimulus three times, participants were asked to repeat the rhythmic sequences they had just heard by pressing the space bar. Their presses of the spacebar were quantized by changing the inter-press times to the nearest interval set (200, 400, 600, 800ms). Then, the ratio of response accuracy was calculated in terms of the presence of hits or rests every 200ms compared to the ones in the target stimuli.
- **Melody Reproduction**: A total of 10 melodies were prepared for the melody reproduction test, each containing a sequence of seven notes (300ms per note). All notes were picked out from a scale of five six-harmonic complex tones (equal amplitude across harmonics) with fundamental frequencies of 220, 246.9, 277.2, 311.1, and 329.6Hz, respectively, corresponding to the first five notes of the A major scale. The melodies were created as follows. First, they all began on the third tone (277.2Hz) of the scale. Then, the next note was the closest note to the previous one on the scale, being either one note higher (246.9Hz) or lower (311.1Hz). This pattern was repeated until all seven notes had been randomly selected. As 220 and 329.6Hz were the lower and upper limits of the melodies, respectively, once one tone reached these two limits, the next note would be either one note closer to the center of the scale or the same as the previous one. Participants listened to the melodies (three times per melody) and were asked to reproduce them by clicking five buttons labeled from “5” to “1” (from the highest tone to the lowest), stretching in a line from the top (“5”) to the bottom (“1”) of the screen. The first seven button presses were recorded and compared to the original melody, and then the mean accuracy ratio was calculated across all ten melodies.

Auditory acuity (formant, pitch, duration). Following the formants developed and detailed in Kachlicka et al. (2019), the discrimination

tests consisted of three sub-tests designed to measure the ability to perceive the spectral and temporal details of a sound, including formant, pitch, and duration discrimination thresholds. Formant and pitch thresholds were used to reveal participants' spectral acuity, while duration thresholds reflected their temporal acuity. For each sub-test, one-hundred continuous synthesized stimuli were generated using custom MATLAB scripts that differed either in terms of formant frequency (F2 = 1500-1700Hz), pitch (F0 = 330-360Hz), or duration (250-500ms) (see Supporting Information-C for a detailed summary). In each test trial, three different stimuli were presented with an inter-stimulus interval of 0.5s. Participants were asked to choose the tone they thought was different from the other two by clicking the corresponding keyboard labeled with "1" or "3" on the screen.

Using the adaptive threshold procedure proposed by Levitt (1971), the difficulty of the task, (i.e., the size of the sound difference), varied according to participants' ongoing performance, which increased after three consecutive correct responses or decreased after an incorrect response. The test terminated after either 70 trials were made or eight reversals (i.e., a difficulty decrease followed by an increase). The final score, ranging from 0 to 100 points, was calculated by averaging the levels of the reversals, beginning from the third level. This score indicates the smallest difference between stimuli that participants were able to discriminate, i.e., the discrimination threshold. Note that lower scores indicate better performance. For example, a result of 10 out of 100 indicates that the smallest difference that a participant could hear was 20Hz for the formant test, 3Hz for pitch test, and 25ms for the duration test.

Reliability. According to the recent test-retest reliability study (Saito & Tierney, forthcoming), the inter-class correlations (ICC) could be considered as good for acuity [$ICC_{(2,2)}$ =.625] and excellent for integration [$ICC_{(2,2)}$ =.843] especially when all the sub-test scores were averaged across (formant, pitch & duration for acuity; rhythm & melody for integration).

RESULTS

The first objective of the statistical analyses was to examine to what degree the experimental group improved thanks to shadowing training via mean-based analyses (ANOVAs). The second objective of the statistical analyses was to explore the extent to which participants' improvements could be related to their auditory processing profiles via variation-based analyses (partial correlations).

Effects of Shadowing Training on Comprehensibility and Accentedness

The results of the descriptive statistics are summarized in Table 1. According to the results of normality tests (Kolmogorov-Smirnov), the rating scores did not significantly differ from a normal distribution ($p > .05$). The results of independent *t*-tests on the pre-test scores (with the alpha level being set at $p < .05$ and adjusted to $p < .025$ via Bonferroni correction) showed that the experimental and control groups' performance was comparable for most of the contexts: comprehensibility in the spontaneous task ($t(45) = -0.149$, $p = .882$, $d = .055$), accentedness in the controlled task ($t(45) = 1.294$, $p = .202$, $d = .499$), and accentedness in the spontaneous task ($t(45) = 0.303$, $p = .763$, $d = .120$). However, the experimental group's comprehensibility score was higher (more comprehensible) than the control group in the controlled task; and the group difference was marginally significant ($t(45) = 2.271$, $p = .028$, $d = .920$). Individual performance as per group was visually plotted in Figure 2. Because there was some form of pre-existing difference (one out of four contexts), the following statistical analyses focused on the extent to which the experimental and control groups improved their proficiency scores throughout the project (Time) but not on how the two groups differed at the time of the post-tests (Group).

To examine the effects of the shadowing training on comprehensibility, we carried out a repeated measures ANOVA with Group as a between-subjects factor and Task and Time as within-subjects factors. The results yielded significant main effects for Time, $F(1, 45) = 108.051$, $p < .001$, $\eta^2 = .706$, Task, $F(1, 45) = 67.318$, $p < .001$, $\eta^2 = .599$, and Group, $F(1, 45) = 21.131$, $p < .001$, $\eta^2 = .987$; and three-way interaction effects for Group ×Task × Time, $F(1, 45) = 5.561$, $p = .023$, $\eta^2 = .110$. According to the results of multiple comparison analysis (with an alpha level $p < .012$ Bonferroni corrected), the experimental group significantly improved their comprehensibility scores over time on both the controlled task ($t = -16.156$, $p < .001$, $d = 2.274$) and the spontaneous task ($t = -9.901$, $p < .001$, $d = 1.662$). Yet, the control group's gains did not reach statistical significance in the controlled task ($t = -0.408$, $p = .693$, $d = 0.309$) and the spontaneous task ($t = -1.646$, $p = .134$, $d = 0.323$).

As for accentedness, the ANOVA revealed significant main effects for Time, $F(1, 45) = 45.387$, $p < .001$, $\eta^2 = .502$, Task, $F(1, 45) = 23.966$, $p < .001$, $\eta^2 = .348$, and Group, $F(1, 45) = 5.916$, $p = .019$, $\eta^2 = .116$; and three-way interaction effects for Group × Task × Time, $F(1, 45) = 18.418$, $p < .001$, $\eta^2 = .290$. Multiple comparison analysis revealed a

TABLE 1
Descriptive Summary of L2 Comprehensibility and Accentedness Ratings at Pre- and Post-Tests

			A. Pre-test				B. Post-test			
					95% CI				95% CI	
			M	*SD*	*Lower*	*Upper*	*M*	*SD*	*Lower*	*Upper*
Experimental (n = 37)	Comprehensibility	Controlled	4.63	0.59	4.43	4.82	5.89	0.53	5.72	6.07
		Spontaneous	3.75	0.53	3.57	3.93	4.58	0.45	4.43	4.73
	Accentedness	Controlled	3.38	0.76	3.12	3.63	4.28	0.75	4.03	4.53
		Spontaneous	2.96	0.54	2.78	3.14	3.50	0.52	3.32	3.67
Control (n = 10)	Comprehensibility	Controlled	4.18	0.36	3.92	4.44	4.22	0.55	3.82	4.62
		Spontaneous	3.78	0.56	3.38	4.18	3.96	0.69	3.47	4.45
	Accentedness	Controlled	3.04	0.59	2.61	3.47	3.20	0.60	2.77	3.63
		Spontaneous	2.90	0.45	2.57	3.23	3.06	0.59	2.64	3.48

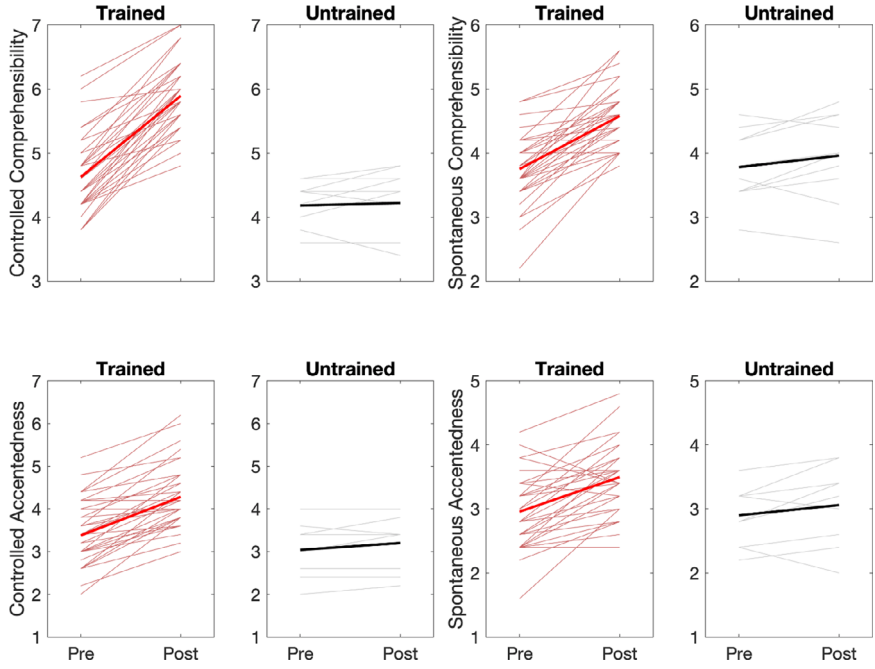

FIGURE 2. Individual performance of comprehensibility and accentedness on controlled and spontaneous tasks over time.

medium effect of improvement for the experimental group on both the controlled and spontaneous tasks, $t = -10.271, -6.945$, $p < .001$, $d = 1.193, 1.020$. By comparison, the control group's performance did not significantly change over time, $t = -2.228, -1.500$, $p = .053, .168$, $d = 0.268, 0.304$.

Individual Differences in Auditory Processing and Instructional Effectiveness

Next, we examined the extent to which the experimental group's instructional gains were associated with their auditory processing profiles. The participants' raw scores are presented in Table 2. Auditory acuity scores indicate the minimum difference that the participants could hear for three types of basic acoustic information (formant, pitch, and duration). Audio-motor integration scores index how accurately (%) the participants could reproduce novel melodic and rhythmic patterns.

Following the procedures of the previous literature (e.g., Kachlicka et al., 2019), a composite auditory acuity score was computed by

TABLE 2
Summary of Raw Scores of Auditory Processing Abilities

Auditory task type	*M*	*SD*
Formant discrimination	82.9Hz	13.67
Pitch discrimination	6.8Hz	12.86
Duration discrimination	69.8ms	15.99
Melody reproduction	48.3%	15.8
Rhythm reproduction	67.1%	8.4

standardizing and averaging the formant, pitch, and duration discrimination scores. Likewise, a composite integration score was computed by standardizing and averaging the rhythm and melody reproduction scores. The results of normality tests (Kolmogorov-Smirnov) suggested that the scores did not significantly deviate from a normal distribution ($p > .05$). Pearson correlation analysis found that the relationship between auditory acuity and integration was not significant ($r = -.251$, $p = .134$), which suggests that the tests tapped into two different aspects auditory processing ability.

Next, partial correlation analyses were carried out to analyze the relationship between participants' gain scores (post-test scores minus pre-test scores) and their auditory processing profiles while controlling for their pre-test scores. Here, we did not conduct the simple correlations between their raw gain and auditory scores because such relationship could have been influenced by participants' pre-test performance (e.g., those with low pre-test scores likely show larger gains because they have more room for improvement).

Table 3 summarizes the correlation coefficients between participants' gain scores and auditory processing after their pre-test scores were factored out. The strength of the correlation coefficients was

TABLE 3
Correlations Between Participants' Gain Scores and Auditory Processing, with Pre-Test Scores Factored Out

		Auditory acuity		Audio-motor integration	
		r	*p*	*r*	*p*
Comprehensibility	Controlled	.122	.479	.295	.081
	Spontaneous	−.048	.781	.229	.179
Accentedness	Controlled	−.024	.891	.308	.068
	Spontaneous	−.315	.062	.482	.003*

*indicates statistical significance at $p < .025$ (Bonferroni corrected)

interpreted with a reference to Plonsky and Oswald's (2014) field-specific benchmark (r =.60 for large; r =.40 for medium; r =.25 for small). The results showed that gain scores for the most difficult aspect of L2 speech learning (i.e., foreign accent reduction in the spontaneous speech task) were significantly related to individual differences in audio-motor integration. The strength of the correlation was medium (r =.482). Interestingly, although the link between acuity and accentedness did not reach statistical significance (p =.062), the strength of the correlation coefficients was small-to-medium (r = −.315). No significant correlations were found for comprehensibility. The size of the correlation coefficients was substantially small (r =

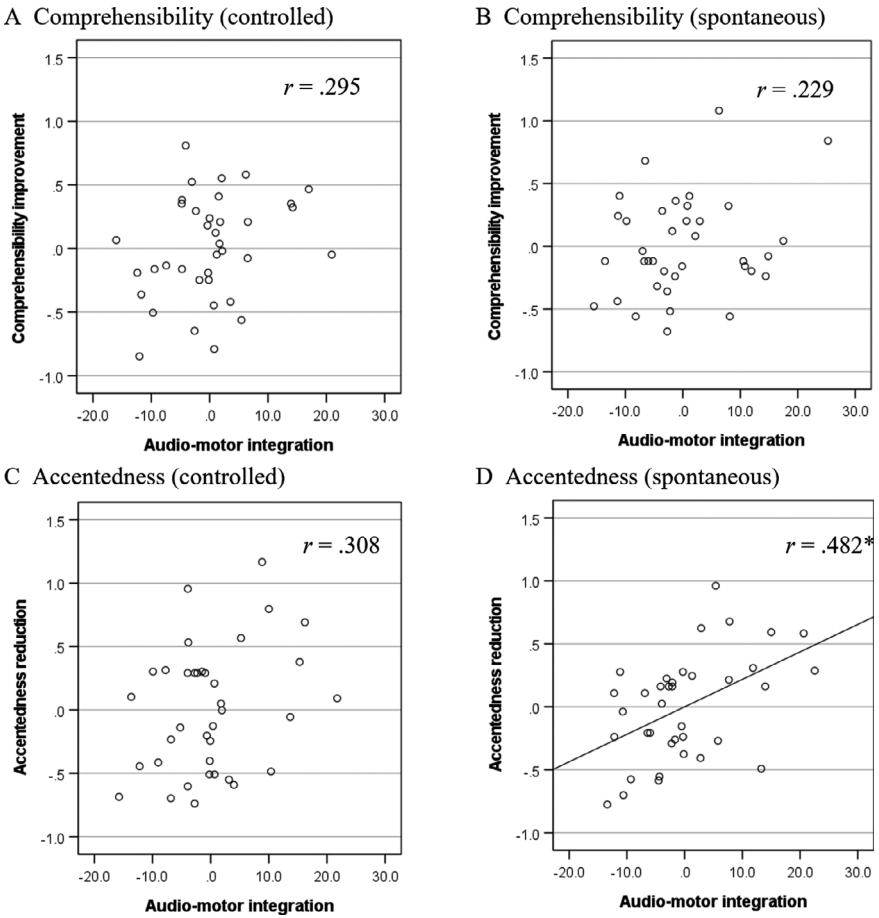

FIGURE 3. Partial correlations between training gains and auditory processing residuals with pre-test scores partialled out. *Note.* Gain scores (−1.0 to 1.5) were centered after participants' pre-existing differences in proficiency (i.e., pre-test scores) were stoically controlled for (i.e., residual scores).

|.048-.295|). The results indicated that participants could enhance comprehensibility regardless of the degree of their ability to integrate auditory patterns and motor sequences. The relationship between participants' change in proficiency and auditory profiles was visually plotted in Figure 3. Note that these gain and auditory scores were residual ones (after participants' pre-test performance was factored out). That is, participants' raw gain scores were centered after participants' pre-existing differences in proficiency (i.e., pre-test scores) were statistically controlled for.

DISCUSSION

In the context of 47 Chinese EFL learners, the current study examined the extent to which mobile-assisted repetition training (shadowing) interacted with auditory processing abilities in the improvement of L2 speech proficiency. Echoing the previous literature (Hamada, 2018), the results showed that shadowing training led to notable gains in comprehensibility (medium-to-large effects) and resulted in some reduction in foreign accentedness (medium effects). The size of gains reported here could be considered as promising given that the provision of explicit instruction is generally small-to-medium, $d = 0.78$ (Saito & Plonsky, 2019), and considering that very few studies have found such a robust impact of instruction on both comprehensibility and accentedness (Saito, 2021). This is arguably ascribed to the fact that the participants in the current study were encouraged to practice shadowing activities via iPad regularly, intensively, and autonomously (Pegrum, 2014). The findings here provide additional support for the pedagogical potential of mobile-based assisted language learning, especially in EFL classrooms, where language exposure is limited in quantity (several hours per week) and quality (the lack of communicatively authentic input) (Muñoz, 2014).

As for the role of individual differences in instructed L2 speech learning, the results of the partial correlation analyses indicated that the degree of speech improvement appeared to be linked to certain auditory processing abilities. Specifically, a significant correlation was found between instructional gains and audio-motor integration ability for accentedness reduction. The findings provide additional longitudinal evidence for the auditory processing hypothesis, which posits that auditory precision serves as a bottleneck for language learning, because it helps learners encode and remember sounds characteristics, helping them make the most of the pedagogical opportunities of each language input (Kachlicka et al., 2019). Thus, higher aptitude learners (i.e., those with more precise auditory acuity or more robust auditory-

motor integration) can attain more advanced L2 proficiency than others even if they engage in the same amount of training (Doughty, 2019).

The results specifically showed that audio-motor integration, rather than auditory acuity, was related to the rate of learning success. This echoes existing cross-sectional evidence showing that advanced L2 learners—and especially those in production-focused EFL classrooms—likely have greater audio-motor integration abilities (rather than auditory acuity) (Saito, Suzukida, et al., 2021). The significant role identified for audio-motor integration over perceptual acuity could be explained by the nature of the training. During the shadowing activities, participants were prompted to focus on repeating what they had heard in the video clips. However, this activity did not require participants to encode the perceptual details of auditory stimuli. Instead, this kind of repetition more closely corresponds to the process of audio-motor integration: Participants needed to track auditory patterns across several seconds, spanning multiple speech segments, and rapidly formulate the appropriate motor sequence for reproducing these patterns. This is exactly the skill tested by the rhythm and melody reproduction tests, which present participants with several seconds of frequency/temporal patterns and ask them to reproduce them. Participants who have difficulty remembering and reproducing acoustic patterns—whether these patterns are drawn from speech or music—may not be able to fully take advantage of shadowing training, either because they struggle to remember acoustic patterns or because they have difficulty translating these patterns into motor output. These reproduction difficulties, then, would result in speech output that is different from the intended speech model, and the comparison between the produced output and the speech target would not facilitate pronunciation learning.

It is interesting that the current study did not find any significant predictive power for auditory acuity in any context. This could be considered as additional evidence that the importance of different auditory processing abilities (acuity vs. integration) varies in accordance with the type of instruction that L2 learners engage in. This is another aspect of individual differences that future studies need to investigate: Certain learning approaches will work better for some individuals, while others would benefit more from alternate approaches. The current findings, for example, suggest that individuals with strong auditory-motor integration skills benefit more from a production-based, shadowing approach, but that individuals with weaker auditory-motor integration might be better off with alternate types of training. Contrastively, research has indicated that auditory acuity may be an important predictor of success in the context of perception-based

training (e.g., Perrachione, Lee, Ha, & Wong, 2011 for high-variability phonetic training). To our knowledge, no empirical studies have investigated the differential effects of auditory acuity and audio-motor integration when L2 learners are engaged in different types of instruction within a single study.

Finally, it needs to be stressed that the aptitude-acquisition link identified here concerned the most difficult instance of L2 speech learning: foreign accentedness reduction at a spontaneous level. In line with Doughty's (2019) aptitude framework, we argue that aptitude is a necessary condition for advanced L2 speech acquisition, which entails successfully learning both easy and difficult L2 speech features. In the current investigation, instructional gains were the greatest when participants speech was elicited using controlled tasks and assessed for comprehensibility. However, aptitude (audio-motor integration) appeared to serve as a crucial deciding factor of nativelike L2 pronunciation attainment when it came to the ability to produce spontaneous speech. It is interesting to note that although the predictive power of auditory acuity failed to reach statistical significance in any contexts (p >.05), the relationship between participants' acuity scores and nativelike pronunciation was considered as small-to-medium (r = −.315; Plonsky & Oswald, 2014). This indicates that different types of aptitude matter when outcome measures concern the relatively difficult aspects of L2 speech learning.

FUTURE DIRECTIONS

With a view toward future replication studies, we would like to raise a number of methodological issues that need to be addressed in the future.

- First and foremost, all the auditory test batteries for acuity (formant, pitch, & duration discrimination) and integration (rhythm & melody reproduction) have been coded in HTML/JavaScript so that everyone (researchers and practitioners alike) can evaluate different types of auditory processing using their own computer. A ZIP file together with a user manual will soon be shared at our in-house website (www.sla-speech-tools.com). For researchers, auditory processing can be adopted as an additional interesting measure of perceptual-cognitive individual differences. For practitioners, the auditory processing profiles of students can be used as one suggestion regarding what type of speech training they can most benefit from (acuity for perception-based training; integration for production-based

training). For more discussion on the scholarly and pedagogical implications of auditory processing, see Saito and Tierney, forthcoming.

- The sustainability of the link between auditory processing and shadowing effectiveness should be measured not only via immediate but also via delayed post-tests.
- Notably, the sentences in the pre- and post-tests were extracted from the materials provided to the experimental group. This methodological feature was considered as a limitation as it may have affected the experimental group's improvement. Future studies should explore the extent to which L2 learners can generalize what they have learned from training to novel sentence contexts (but see the results of the spontaneous production tasks in the Results section).
- The findings of the current study should be replicated with a larger sample size and an equal number of participants per group.
- Although the current study found a significant role for auditory processing in instructed L2 speech learning, we acknowledge that the behavioral tasks adopted in the current study inevitably tap into various aspects of cognitive abilities beyond audio-motor integration and auditory acuity, e.g., attentional control and auditory short-term memory (McArthur & Bishop, 2005). In order to tease out the effects of auditory processing in L2 speech acquisition, future studies may need to assess and factor out participants' individual differences in executive functioning (cf. Saito et al., in press).
- The current study (together with precursor research such as Saito, Suzukida, et al., 2021) has suggested that auditory acuity and audio-motor integration represent two different kinds of auditory processing abilities. Extending our above-mentioned discussion, we could argue that the discrimination and memory tests may differ not only on the role of motor output but also on the time scales of the patterns to be detected (precise characteristics of single sounds versus overall pattern across multiple sounds). This could explain why audio-motor integration predicted the rate of success in shadowing training wherein participants were encouraged to track and repeat the broad gist of the overall pattern than to be able to encode precise details of single sounds.
- To further develop a more fine-grained framework of auditory processing, future studies should test how different types of auditory processing can be uniquely tied to L2 speech acquisition when learners receive different types of auditory input. For example, whereas previous studies have mostly examined the role

of auditory processing in language-focused training (e.g., high-variability phonetic training and shadowing), few scholars have ever delved into how auditory processing can facilitate L2 speech acquisition when participants receive more communicatively oriented instruction (e.g., recasts; Lee & Lyster, 2016; Saito, 2013b). Such studies could reveal whether, to what degree and how auditory processing (acuity vs. audio-motor integration) can be predictive of L2 speech learning on explicit *and* implicit modes.

- In the cognitive psychology literature, audio-motor integration ability has been linked to music-training experience (e.g., Tierney et al., 2017). Future studies should survey participants' backgrounds in music training and their impact on auditory processing and L2 speech learning (cf. Zheng et al., 2021).
- Given evidence that focused training can help improve auditory processing ability (Hayes, Warrier, Nicol, Zecker, & Kraus, 2003 for 35-40 hours of commercial auditory processing training programs), it would be intriguing to examine whether the effectiveness of instructed L2 speech learning can be boosted if pronunciation and auditory skills are improved simultaneously. It can be hypothesized that such a combined approach may help optimize the process of L2 speech learning and can lead to the acquisition of advanced-level L2 speech proficiency (cf., see Hayashi, 2019 for the role of working memory training in L2 grammar learning).
- The current study highlights the relationship between the participants' auditory processing and L2 speech learning on a broad level. However, it remains unknown precisely how auditory processing facilitates L2 speech learning at a fine-grained level. Although auditory integration abilities were linked to L2 accentedness in the current study, it remains unclear how auditory integration (rather than acuity) could help participants reduce their degree of accentedness. To remedy this weakness, future studies should look at L2 speech acquisition at a more fine-grained level, i.e., by examining specific dimensions of L2 speech acquisition (e.g., second and third formants and duration for English [r] and [l] by Japanese learners).

REFERENCES

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Appel, R., Trofimovich, P., Saito, K., Isaacs, T., & Webb, S. (2019). Lexical aspects of comprehensibility and nativeness from the perspective of native-speaking

English raters. *ITL-International Journal of Applied Linguistics*, *170*(1), 24–52. https://doi.org/10.1075/itl.17026.app
Boets, B., Vandermosten, M., Poelmans, H., Luts, H., Wouters, J., & Ghesquiere, P. (2011). Preschool impairments in auditory processing and speech perception uniquely predict future reading problems. *Research in Developmental Disabilities*, *32*(2), 560–570. https://doi.org/10.1016/j.ridd.2010.12.020
Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching pronunciation hardback with audio CDs (2): A course book and reference guide*. Cambridge University Press.
Council of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*(2), 218–236. https://doi.org/10.1016/0749-596X(92)90012-M
Darcy, I., Park, H., & Yang, C. L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences*, *40*, 63–72. https://doi.org/10.1016/j.lindif.2015.04.005
De Pijper, J. R., & Sanderman, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *The Journal of the Acoustical Society of America*, *96*(4), 2037–2047. https://doi.org/10.1121/1.410145
Derwing, T. (2003). What do ESL students say about their accents? *Canadian Modern Language Review*, *59*(4), 547–567. https://doi.org/10.3138/cmlr.59.4.547
Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A 7-year study. *Language Learning*, *63*(2), 163–185. https://doi.org/10.1111/lang.12000
Derwing, T. M., Munro, M. J., Foote, J. A., Waugh, E., & Fleming, J. (2014). Opening the window on comprehensible pronunciation after 19 years: A workplace training study. *Language Learning*, *64*(3), 526–548. https://doi.org/10.1111/lang.12053
Dormann, C., & Guthier, C. (2019). Successful and positive learning through study crafting: A self-control perspective. In *Frontiers and advances in Positive Learning in the Age of InformaTiOn (PLATO)* (pp. 57–72). Springer, Cham. https://doi.org/10.1007/978-3-030-26578-6_5
Doughty, C. J. (2019). Cognitive language aptitude. *Language Learning*, *69*, 101–126. https://doi.org/10.1111/lang.12322
EIKEN Foundation of Japan. (2016). *EIKEN pre-1 level: Complete questions collection*. Tokyo: Oubunsha.
Flege, J., & Bohn, O.-S. (2021). The revised speech learning model. In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress*. Cambridge: Cambridge University Press.
Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*(5), 3125–3134. https://doi.org/10.1121/1.413041
Goswami, U. (2015). Sensory theories of developmental dyslexia: Three challenges for research. *Nature Reviews Neuroscience*, *16*(1), 43–54. https://doi.org/10.1038/nrn3836
Hamada, Y. (2018). Shadowing for pronunciation development: Haptic-shadowing and IPA-shadowing. *Journal of Asia TEFL*, *15*(1), 167. https://doi.org/10.18823/asiatefl.2018.15.1.11.167

Hayashi, Y. (2019). Investigating effects of working memory training on foreign language development. *The Modern Language Journal*, *103*, 665–685. https://doi.org/10.1111/modl.12584

Hayes, E. A., Warrier, C. M., Nicol, T. G., Zecker, S. G., & Kraus, N. (2003). Neural plasticity following auditory training in children with learning problems. *Clinical Neurophysiology*, *114*(4), 673–684. https://doi.org/10.1016/S1388-2457(02)00414-5

He, Q., Xue, G., Chen, C., Chen, C., Lu, Z. L., & Dong, Q. (2013). Decoding the neuroanatomical basis of reading ability: A multivoxel morphometric study. *Journal of Neuroscience*, *33*(31), 12835–12843. https://doi.org/10.1523/JNEUROSCI.0449-13.2013

Hornickel, J., & Kraus, N. (2013). Unstable representation of sound: A biological marker of dyslexia. *Journal of Neuroscience*, *33*(8), 3500–3504. https://doi.org/10.1523/JNEUROSCI.4205-12.2013

Hsieh, K. T., Dong, D. H., & Wang, L. Y. (2013). A preliminary study of applying shadowing technique to English intonation instruction. *Taiwan Journal of Linguistics*, *11*(2), 43–65. https://doi.org/10.6519/TJL.2013.11(2).2

Isaacs, T. (2008). Towards defining a valid assessment criterion of pronunciation proficiency in non-native English-speaking graduate students. *Canadian Modern Language Review*, *64*(4), 555–580. https://doi.org/10.3138/cmlr.64.4.555

Isaacs, T., Trofimovich, P., & Foote, J. A. (2018). Developing a user-oriented second language comprehensibility scale for English-medium universities. *Language Testing*, *35*(2), 193–216. https://doi.org/10.1177/0265532217703433

Kachlicka, M., Saito, K., & Tierney, A. (2019). Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound. *Brain and Language*, *192*, 15–24. https://doi.org/10.1016/j.bandl.2019.02.004

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, *94*(4), 554–566. https://doi.org/10.1111/j.1540-4781.2010.01091.x

Kidd, G. R., Watson, C. S., & Gygi, B. (2007). Individual differences in auditory abilities. *The Journal of the Acoustical Society of America*, *122*(1), 418–435. https://doi.org/10.1121/1.2743154

Lee, A. H., & Lyster, R. (2016). The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition*, *38*(1), 35–64. https://doi.org/10.1017/S0272263115000194

Levis, J. M. (2018). *Intelligibility, oral communication, and the teaching of pronunciation.* Cambridge University Press.

Levitt, H. C. C. H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2B), 467–477. https://doi.org/10.1121/1.1912375

Li, M., & DeKeyser, R. (2017). Perception practice, production practice, and musical ability in L2 Mandarin tone-word learning. *Studies in Second Language Acquisition*, *39*(4), 593–620. https://doi.org/10.1017/S0272263116000358

Lord, G. (2015). "I don't know how to use words in Spanish": "Rosetta Stone" and learner proficiency outcomes. *The Modern Language Journal*, *99*(2), 401–405. http://www.jstor.org/stable/43650036

Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P., & Lee, C. S. (1992). Prosodic effects in minimal attachment. *The Quarterly Journal of Experimental Psychology*, *45*(1), 73–87. https://doi.org/10.1080/14640749208401316

Martinsen, R., Montgomery, C., & Willardson, V. (2017). The effectiveness of video-based shadowing and tracking pronunciation exercises for foreign language learners. *Foreign Language Annals*, *50*(4), 661–680. https://doi.org/10.1111/flan.12306

McArthur, G. M., & Bishop, D. V. (2005). Speech and non-speech processing in people with specific language impairment: A behavioural and electrophysiological study. *Brain and Language*, *94*(3), 260–273. https://doi.org/10.1016/j.bandl.2005.01.002

Moore, B. C. (2012). *An introduction to the psychology of hearing*. Brill.

Mueller, J. L., Friederici, A. D., & Männel, C. (2012). Auditory perception at the root of language learning. *Proceedings of the National Academy of Sciences*, *109*(39), 15953–15958. https://doi.org/10.1073/pnas.1204319109

Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics*, *35*, 463–482. https://doi.org/10.1093/applin/amu024

Nagle, C. (2018). Motivation, comprehensibility, and accentedness in L2 Spanish: Investigating motivation as a time-varying predictor of pronunciation development. *The Modern Language Journal*, *102*(1), 199–217. https://doi.org/10.1111/modl.12461

Nagle, C. L. (2021). Assessing the state of the art in longitudinal L2 pronunciation research: Trends and future directions. *Journal of Second Language Pronunciation*, *7*, 154–182. https://doi.org/10.1075/jslp.20059.nag

Ortega, L., & Iberri-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics*, *25*, 26–45. https://doi.org/10.1017/S0267190505000024

Pegrum, M. (2014). *Mobile learning: Languages, literacies and cultures*. Springer.

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–472. https://doi.org/10.1121/1.3593366

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, *2*(4), 411–440. https://doi.org/10.2307/40285311

Qin, Z., Zhang, C., & Wang, W. S. Y. (2021). The effect of Mandarin listeners' musical and pitch aptitude on perceptual learning of Cantonese level-tones. *The Journal of the Acoustical Society of America*, *149*(1), 435–446. https://doi.org/10.1121/10.0003330

Saito, K. (2013a). Age effects on late bilingualism: The production development of/ɹ/by high-proficiency Japanese learners of English. *Journal of Memory and Language*, *69*(4), 546–562. https://doi.org/10.1016/j.jml.2013.07.003

Saito, K. (2013b). The acquisitional value of recasts in instructed second language speech learning: Teaching the perception and production of English /ɹ/ to adult Japanese learners. *Language Learning*, *63*(3), 499–529. https://doi.org/10.1111/lang.12015

Saito, K. (2021). What characterizes comprehensible and native-like pronunciation among English-as-a-Second-Language speakers? Meta-analyses of phonological, rater, and instructional factors. *TESOL Quarterly*, *55*, 866–900. https://doi.org/10.1002/tesq.3027

Saito, K., Haining, C., Suzukida, Y., Yuichi, S., Jeong, H., Revesz, A., Sugiura, M., & Tierney, A. (in press). Does domain-general auditory processing uniquely

explain the outcomes of second language speech acquisition, even once cognitive and demographic variables are accounted for? *Bilingualism: Language and Cognition*.

Saito, K., Kachlicka, M., Sun, H., & Tierney, A. (2020). Domain-general auditory processing as an anchor of post-pubertal second language pronunciation learning: Behavioural and neurophysiological investigations of perceptual acuity, age, experience, development, and attainment. *Journal of Memory and Language*, *115*, 104168. https://doi.org/10.1016/j.jml.2020.104168.

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, *69*(3), 652–708. https://doi.org/10.1111/lang.12345

Saito, K., Sun, H., Kachlicka, M., Alayo, J. R. C., Nakata, T., & Tierney, A. (2020). Domain-general auditory processing explains multiple dimensions of L2 acquisition in adulthood. *Studies in Second Language Acquisition*, 1–30, https://doi.org/10.1017/S0272263120000467

Saito, K., Suzuki, S., Oyama, T., & Akiyama, Y. (2021). How does longitudinal interaction promote second language speech learning? Roles of learner experience and proficiency levels. *Second Language Research*, *37*, 547–571. https://doi.org/10.1177/0267658319884981

Saito, K., Suzukida, Y., Tran, M., & Tierney, A. (2021). Domain-general auditory processing partially explains second language speech learning in classroom settings: A review and generalization study. *Language Learning*. https://doi.org/10.1111/lang.12447

Saito, K., & Tierney, A. (forthcoming). Domain-general auditory processing as a conceptual and measurement framework for second language speech learning aptitude: Systematic review and test development.

Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*(4), 439–462. https://doi.org/10.1093/applin/amv047

Saito, K., & van Poeteren, K. (2012). Pronunciation-specific adjustment strategies for intelligibility in L2 teacher talk: Results and implications of a questionnaire study. *Language Awareness*, *21*(4), 369–385. https://doi.org/10.1080/09658416.2011.643891

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*(1), 187. https://doi.org/10.1017/S0142716417000418

Sun, H., Saito, K., & Tierney, A. (2021). A longitudinal investigation of explicit and implicit auditory processing in L2 segmental and suprasegmental acquisition. *Studies in Second Language Acquisition*, 1–23, https://doi.org/10.1017/S0272263120000649

Suzuki, Y. (2021). Individual differences in memory predict changes in breakdown and repair fluency but not speed fluency: A short-term fluency training intervention study. *Applied Psycholinguistics*, 1–27, https://doi.org/10.1017/S0142716421000187

Tierney, A., White-Schwoch, T., MacLean, J., & Kraus, N. (2017). Individual differences in rhythm skills: Links with neural consistency and linguistic ability. *Journal of Cognitive Neuroscience*, *29*(5), 855–868. https://doi.org/10.1162/jocn_a_01092

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*(4), 905–916. https://doi.org/10.1017/S1366728912000168

Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2016). Flawed self-assessment: Investigating self-and other-perception of second language speech. *Bilingualism: Language and Cognition*, *19*(1), 122–140. https://doi.org/10.1017/S1366728914000832

Venkatagiri, H. S., & Levis, J. M. (2007). Phonological awareness and speech comprehensibility: An exploratory study. *Language Awareness*, *16*(4), 263–277. https://doi.org/10.2167/la417.0

Wang, L., Li, X., & Yang, Y. (2014). A review on the cognitive function of information structure during language comprehension. *Cognitive Neurodynamics*, *8*(5), 353–361. https://doi.org/10.1007/s11571-014-9305-1

Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, *50*(1), 509–535. https://doi.org/10.1146/annurev.psych.50.1.509

White-Schwoch, T., Woodruff Carr, K., Thompson, E. C., Anderson, S., Nicol, T., Bradlow, A. R., … Kraus, N. (2015). Auditory processing in noise: A preschool biomarker for literacy. *PLoS Biology*, *13*(7), e1002196. https://doi.org/10.1371/journal.pbio.1002196

Zheng, C., Saito, K., & Tierney, A. (2021). Successful second language pronunciation learning is linked to domain-general auditory processing rather than music aptitude. *Second Language Research*. https://doi.org/10.1177/0267658320978493

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Supplementary Material